

AI's philosophical underpinnings: A thinking person's walk through the twists and turns of artificial intelligence's meandering path

Silvano Colombano

Few human endeavors can be viewed both as extremely successful and unsuccessful at the same time. This is typically the case when goals have not been well defined or have been shifting in time. This has certainly been true of Artificial Intelligence (AI).

The behaviorist model

The nature of intelligence has been the object of much thought and speculation throughout the history of philosophy. It is in the nature of philosophy that real headway is sometimes made only when appropriate tools become available. For instance, the nature and behavior of physical objects was a major topic of philosophy. That is until the experimental method and the advent of calculus allowed for the development of Physics.

Similarly the computer, coupled with the ability to program (at least in principle) any function, appeared to be the tool that could tackle the notion of intelligence. To suit the tool, the problem of the “nature” of intelligence was soon sidestepped in favor of this notion: If a probing conversation with a computer could not be distinguished from a conversation with a human, then “artificial” intelligence had been achieved. This notion became known as the “Turing test”, after the mathematician Alan Turing who proposed it in 1950.

This challenge quickly attracted the best computer scientists in a worldwide search for techniques and principles of what soon became known as the field of Artificial Intelligence. The early efforts focused on creating “general problem solvers” like, for instance, the Soar system (Newell, Laird and Rosenbloom) which attempted to solve problems by breaking them down into sub-goals.

Conceptually rich and interesting, these early efforts gave rise to a large portion of the field’s framework. Key to artificial intelligence, rather than the “number crunching” typical of computers until then, was viewed as the ability to manipulate symbols and make logical inferences. To facilitate these tasks, “AI languages” such as LISP and Prolog were invented and used widely in the field.

That this quest never strayed far from rigorous mathematical underpinnings was both its strength and its limitation. Its strength was to open a new fertile area of computer science. Its limitation was that “real world” problems tended to be too complex for the limitations imposed by mathematical rigor and the constraints of logic and symbol manipulation. Therefore, much effort continued to be focused on “toy problems.”

One idea that emerged and enabled some success with real world problems was the notion that “most” intelligence really resided in knowledge. A phrase attributed to Feigenbaum, one of the pioneers, was “knowledge is the power.”

With this premise, the problem is shifted from “how do we solve problems” to “how do we represent knowledge.” A good knowledge representation scheme could allow one to draw conclusions from given premises. Such schemes took forms such as *rules, frames and scripts*. It allowed the building of what became known as “expert systems” or “knowledge based systems” (KBS).

These types of systems could indeed help in real world problems (the author led a project for the first expert system to aid astronauts in performing some scientific experiments. It was called PI-in-a-Box). The technology that ensued from expert systems gave rise to the first instance of an “AI industry.” Consulting “Knowledge Engineers” and products (Shells) could take some of the drudgery out of building these types of systems.

The enthusiasm of this time, however, masked an important shift that had been made by this technology: “Real world” solutions were obtained by keeping the system’s focus extremely narrow and limited in scope. These systems were, and, to a large extent, remain extremely “fragile.” That is, unexpected inputs or straying from the scope of the system could easily result in unexpected and erroneous results. The most difficult aspects of intelligence to incorporate appeared to be understanding a) one’s limits of knowledge and b) the, unfortunately, elusive “common sense.”

The very usefulness and continuing success of these types of systems has also brought to light the fundamental limitation of the behaviorist model of intelligence. This model has difficulty coping with the fact that intelligence seems to reside in the ability to achieve one’s expertise and to use it appropriately more than, or certainly in addition to, the expertise itself.

Again, this realization shouldn’t take away from the continuing improvements and successes in these types of systems. Model Based Reasoning has emerged as a powerful approach to diagnosis, and planning and scheduling systems have had much success as well. The point is that AI, now increasingly called “Symbolic AI,” has produced a new branch of computer science. Along with it, powerful tools have been created for knowledge representation, symbol manipulation, searching and optimization. AI is alive and well. However, many opine that its picture of intelligence is too fragmented to represent a satisfactory model of cognition.

News of its death have been greatly exaggerated

Enter neural networks. They were dismissed early on for lack of computational power by Minsky and Papert (quite correctly, given the extreme simplicity of the early “Perceptron” model). However, neural networks have re-emerged as the solution to at least some problems that have dogged symbolic AI. Based on a highly abstracted notion of “neurons,” connections and “synaptic weights,” neural networks can be presented with input-output sets and can learn to perform very complex mapping functions.

These systems’ power comes from side stepping the problem of finding good knowledge representation and inferencing rules. Neural networks can be thought of as building “implicit” models of systems in the form of synaptic weights. Also, they are inherently capable of learning by simply modifying these weights. Major areas of application have been classification and pattern recognition problems. But neural algorithms are being adapted, with varying degrees of success, to all types of “traditional” AI problems.

Neural networks’ lure is threefold: 1) adaptation to changes (a fundamental characteristic of real world situations) is a natural offshoot of the learning algorithms used to “train” these systems; 2) results tend to be robust with respect to input variations, and 3) some comfort can be derived from the fact that the fundamental governing principles, if not the actual algorithms, have some strong analogies with how parts of the biological nervous system works.

With this paradigm shift in the approach to artificial intelligence there has also been a sociological shift. Neural network research and industry is dominated by engineers and physicists rather than by computer scientists. The tendency has been to view intelligent processes as closely coupled to the hardware.

The name “sub-symbolic” is often being used for this approach to intelligence. An attempt has also been made to re-define the field by using the name “Computational Intelligence.” The term represents an approach that combines neural networks, fuzzy logic (which can be shown to be very compatible with neural networks) and genetic algorithms (more on these).

Is this an improvement?

We certainly have increased the choice of tools we have for building systems that display intelligent behavior. But do neural network architectures give a greater insight into intelligence than the programs and data structures of symbolic AI? Are we closer to the kind of system that can achieve expertise and understands its own limitations? In some sense, both symbolic AI and sub-symbolic AI are victims of their success.

Once it became obvious that real world problems could be solved, both camps became fragmented and oriented towards accomplishing specific tasks by any means available. In both cases, the quest for understanding the nature of intelligence has been largely “lost in the shuffle.”

It is true that the symbolic sub-symbolic dichotomy is crying out for a new integrated view. Perhaps integration attempts will bear fruit. Perhaps we have simply reached two dead ends and some rethinking is necessary.

So what is “intelligence”?

Another indication that we may have lost sight of our quest is that more and more systems are being referred to as intelligent or “smart.” We have *intelligent control*, *intelligent design*, *smart bombs*, *smart appliances*, *intelligent agents* and so forth. Everything that has some ability to “decide,” even on the basis of very simple rules, is being termed “intelligent.” We appear to have really defined the problem away. At the same time, a new insight may be emerging in the fact that intelligence is a much more pervasive concept than the one implied by the Turing test.

Intelligence and evolution

Is intelligence in the product or in the process? A car engine wouldn’t be viewed as intelligent, but the process that produced it certainly is. At the same time, a talking dashboard that reminds us to hook up our seatbelt would probably be called “intelligent.” Even though, it might not consist of much more than a glorified switch.

A biological cell hasn’t been viewed as an example of intelligent behavior. Yet, if we could build a machine with similar properties, capable to change its state in response to environmental conditions, capable of self-repair and even self-reproduction, current definitions of intelligence would certainly apply. At the very least we would view the process that created it as intelligent.

If the essence of intelligence is an ability to solve problems, then evolution is the ultimate intelligent process. It has the extraordinary ability to solve the survival problem and to construct the myriad of systems that each living entity needs to survive. These include growth, reproduction, digestion, sensing and, ultimately, the seat of the type of intelligence we have really been after, the brain itself.

An obvious difference between the problem solving mechanisms of evolution (natural selection) and those of the brain (neuronal activity) is the time scale for these processes. Fortunately, computer implementations of natural selection, like evolutionary computation in its various forms such as genetic algorithms and genetic programming,

are making this a mute point for many applications. For example, we can now use evolutionary computation to produce analog circuit designs. In fact, we cannot tell them apart from those “produced” by the human brain.

How far can we push this? Can we finally understand human intelligence by evolving it? Maybe...but the important result of these developments is a different one: we can no longer push the problem away by re-defining it.

If we want to produce solutions to difficult engineering problems, we have now developed a large number of techniques we can use and refine. Posterity will decide whether the term “intelligence” is appropriate either for the techniques or for the solutions.

In any case, much exciting research and engineering work lies ahead. If we truly want to understand and ultimately produce human-like intelligence, we may have to return to cognitive models grounded in the neural “hardware” of the brain. Symbolic, sub-symbolic and evolutionary computing will certainly have roles to play in building these models. But, in the opinion of this author, the common framework will have to come from brain architecture. Abstracting relevant principles remains extremely difficult until we really understand the interplay between the sensing level and the highest levels of cognition.

Read more about it

- W. Clancey, S. Smoliar and M. Sefik eds. *Contemplating Minds: a Forum for Artificial Intelligence*. MIT Press 1994.
- D. Crevier, *AI: the Tumultuous History of the Search for Artificial Intelligence*, Harper Collins 1993.
- P. Flach and R. Meersman eds. *Future Directions in Artificial Intelligence*, North Holland 1991.
- D. Fogel, *Evolutionary Computation*, IEEE Press 1995.
- H. Mishkoff, *Understanding Artificial Intelligence*, Howard Sams & Co. 1988.

About the author

Dr. Silvano Colombano received an M.A. in Physics and a Ph.D. in Biophysical Sciences from the State University of New York at Buffalo (1977). He has spent most of his working career at NASA-Ames Research Center first as a researcher in Closed Ecological Life Support Systems (CELSS) and later in Artificial Intelligence. He began the development work on the Astronaut Science Advisor (a.k.a. PI-in-a-box) and managed the project until its deployment on SLS-2 (Space Shuttle STS -58) in 1993. He has been doing research and development work in Artificial Neural Networks, Genetic Algorithms and Fuzzy Logic, and has taught courses, in these areas for the University of California at Berkeley, extended education. He has over 30 technical publications and now leads the Biologically Inspired Optimization Systems (BIOS) Group in the Computational Sciences Division at NASA-Ames.